

InCoder: A Generative Model for Code Infilling and Synthesis

Year: 2022 | Citations: 781 | Authors: Daniel Fried, Armen Aghajanyan, Jessy Lin, Sida I. Wang, Eric Wallace

Abstract

Code is seldom written in a single left-to-right pass and is instead repeatedly edited and refined. We introduce InCoder, a unified generative model that can perform program synthesis (via left-to-right generation) as well as editing (via infilling). InCoder is trained to generate code files from a large corpus of permissively licensed code, where regions of code have been randomly masked and moved to the end of each file, allowing code infilling with bidirectional context. Our model is the first generative model that is able to directly perform zero-shot code infilling, which we evaluate on challenging tasks such as type inference, comment generation, and variable re-naming. We find that the ability to condition on bidirectional context substantially improves performance on these tasks, while still performing comparably on standard program synthesis benchmarks in comparison to left-to-right only models pretrained at similar scale. The InCoder models and code are publicly released.

<https://sites.google.com/view/incoder-code-models>